

College Center for Library Automation Tallahassee, FL

- Susan B. Campbell
(scampbell@cclaflorida.org)
- Jim McGill
(jmcgill@cclaflorida.org)

automating retrieval and reporting of database usage statistics for a consortium

- CCLA provides and maintains the Library Information Network for 28 Community Colleges (LINCC) for Florida's 65+ community college libraries.
- db statistics we're collecting and reporting
 - 19 vendors
 - over 200 databases
 - monthly reports by database, campus, statewide
 - on demand
- customers for monthly reports
 - 28 community colleges in Florida
 - internal reports

automating retrieval and reporting of database usage statistics for a consortium

- **problem**
 - what we were doing and why it doesn't work
- **solution**
 - the pieces, the parts and how they fit together
- **future**
 - what we've learned and our expectations

the problem

- excel excess

Searches Vendor Totals									
	06-07YTD	Dec-07	07-08YTD	Dec-08	06-07YTD	Jan-08	07-08YTD	Jan-07	06-07YTD
ABC-CLIO	11,973	0	47,941	1,532	13,511	0	47,941	0	21,587
Alexander Street	0	0	2,176	0	291	0	0	0	0
AR Tator	0	0	0	0	0	0	0	194	487
CO Press (CO Researcher)	265	155	0	0	0	0	0	0	0
Columbia Univ PR - Grangers	23,844	2,171	28,452	2,230	26,074	0	28,452	0	29,002
eBooks (NetLibrary)	0	0	0	0	0	0	0	0	0
EBSCO (COUNTER DB)	46,812	5,303	63,170	4,404	51,296	0	63,170	0	68,668
EBSCO Statewide DB Total	25,931	2,636	30,263	2,851	31,292	0	30,263	0	36,002
EBSCO Other DB Total	912,219	136,415	1,176,849	73,418	1,079,571	0	1,176,849	0	1,135,568
Facts.com	46,106	7,789	50,159	6,911	57,891	0	50,159	0	58,164
FirstSearch	65,603	9,285	74,813	8,177	83,991	0	74,813	0	81,744
FirstSearch Statewide DB Total	5,210	1,542	6,752	1,400	7,300	0	6,752	0	7,444
FirstSearch Other DB Total	72,598	10,743	83,565	6,771	90,291	0	83,565	0	89,300
Gale Group	69,267	6,196	75,463	6,911	82,374	0	75,463	0	82,374
Gale Group Statewide DB Total	11,631	2,906	14,537	3,055	17,592	0	14,537	0	16,676
Gale Group Other DB Total	1,639,793	162,409	1,802,569	1,856	1,821,782	0	1,802,569	0	1,818,708
Greenwood	1,451,933	148,175	1,600,108	151,402	1,601,510	0	1,600,108	0	1,615,510
Heaven's Premium	178,800	13,934	192,734	13,452	206,186	0	192,734	0	211,138
ISTOR	0	610	610	0	610	0	610	0	610
Microdex PDR	72,615	1,372	73,987	1,372	75,359	0	73,987	0	75,359
NewsBank	0	0	0	0	0	0	0	0	0
Oxford (Hornell Grove)	0	15,265	15,265	0	15,265	0	15,265	0	15,265
ProQuest	59,496	9,977	69,473	9,712	79,185	0	69,473	0	79,185
PreQuest Statewide DB Total	11,798	3,214	15,012	3,086	18,098	0	15,012	0	18,098
PreQuest Other DB Total	213,656	26,061	239,717	24,729	264,446	0	239,717	0	264,446
Wilson	179,625	4,859	184,484	5,036	189,520	0	184,484	0	194,556
Wilson Statewide DB Total	32,807	22,106	54,913	23,219	78,126	0	54,913	0	78,126
Wilson Other DB Total	321,472	56,884	378,356	57,148	436,646	0	378,356	0	436,646
Credoreference (Hornell Grove)	534	55,586	56,120	29,971	86,091	0	56,120	0	86,091
Statewide Subtotal	3,009,332	386,401	3,395,733	280,872	3,676,605	0	3,395,733	0	3,757,273
Grand Total	3,473,431	440,633	3,914,064	322,783	4,236,847	0	3,914,064	0	4,259,417
SW Contracts	0	0	0	0	0	0	0	0	0
CO Press - CO Researcher	23,844	2,171	28,452	2,230	32,283	0	28,452	0	29,002
Columbia Univ PR - Grangers	0	0	0	0	0	0	0	0	0
eBooks	46,812	5,303	63,170	4,404	73,689	0	63,170	0	68,668
EBSCO (COUNTER DB)	866,573	128,626	995,199	69,225	1,064,424	0	995,199	0	1,070,784
Facts.com	69,393	7,743	77,136	7,498	84,634	0	77,136	0	81,744
FirstSearch	66,967	9,910	76,877	6,555	83,432	0	76,877	0	81,744
Greenwood	15,973	2,326	18,299	1,689	20,988	0	18,299	0	21,587
Heaven's Premium	0	1,859	1,859	0	1,859	0	1,859	0	1,859
ISTOR	120,304	0	120,304	0	120,304	0	120,304	0	120,304
Microdex PDR	62,739	0	62,739	0	62,739	0	62,739	0	62,739
NewsBank	0	0	0	0	0	0	0	0	0
Oxford (Hornell Grove)	0	0	0	0	0	0	0	0	0
ProQuest	309,089	309,089	309,089	309,089	309,089	0	309,089	0	309,089
PreQuest Statewide DB Total	120,304	0	120,304	0	120,304	0	120,304	0	120,304
PreQuest Other DB Total	62,739	0	62,739	0	62,739	0	62,739	0	62,739
Wilson	237,531	0	237,531	0	237,531	0	237,531	0	237,531
Wilson Statewide DB Total	0	0	0	0	0	0	0	0	0
Wilson Other DB Total	0	0	0	0	0	0	0	0	0
Credoreference (Hornell Grove)	0	0	0	0	0	0	0	0	0
Statewide Subtotal	0	0	0	0	0	0	0	0	0
Grand Total	0	0	0	0	0	0	0	0	0

March 20, 2008

Electronic Resources and Libraries



the problem

- vendor variety

repeat 28 times or more for each vendor
(and sometimes each database)

Monthly LINCCWeb Reports

L001	LINCCWeb/Catalog Searches	2007-12-31	Pdf	Zip	Description
L002	External Database Connections	2008-01-31	Pdf	Zip	Description
L006	FirstSearch Statistics (RETIRED)	2006-12-31	Pdf	Zip	Description
L009	SFX/Metalib Statistics	2008-01-31	Pdf	Zip	Description
L010	Vendor Statistics - College Searches	2008-01-31	Pdf	Zip	Description
L011	Vendor Statistics - Statewide Searches	2008-01-31	Pdf	Zip	Description
L012	Ebooks Usage - College Title Requests	2008-01-31	Pdf	Zip	Description
L013	Ebooks Usage - Statewide Title Requests	2008-01-31	Pdf	Zip	Description
L014	Metalib Usage Comparison	2008-01-31	Pdf	Zip	Description
L015	SFX Usage Comparison	2008-01-31	Pdf	Zip	Description

GO BACK USER INFORMATION FORGOTTEN PASSWORDS LOG OUT

1 / 9 64.9% Find

L011, Vendor Statistics - Statewide - Searches

Statewide Searches

	Jul-07	Aug-07	Sep-07	Oct-07	Nov-07	Dec-07	Jan-08	Feb-08	Mar-08	Apr-08	May-08	Jun-08	Total
ABC-CLIO	551	955	1015	1088	1597	1525	2957	0	0	0	0	0	51,887
American Government	2	2	2	112	2,847	51	779	0	0	0	0	0	1,720
American History	1	207	2	78	6	30	1	0	0	0	0	0	2,320
World Geography	106	1	169	80	8	52	9	0	0	0	0	0	377
World History: Ancient	1,013	21	113	181	1,117	1,117	1,220	0	0	0	0	0	10,000
World History: Modern	51	73	1,605	2	1	1,022	356	0	0	0	0	0	10,000
Understanding Controversy & Society						3	0	0	0	0	0	0	1
AccuWeather													
AP Multimedia Archive ^{AB}													
Alexander Street	112	108	281	51	351	15	311	0	0	0	0	0	13
African American Song ^{A3}	2	2	2	4	4			0	0	0	0	0	1
Classical Music Library ^{A3}	1	1	1	14	1	3		0	0	0	0	0	12
Contemporary World Music ^{A3}	3	3	3	3	3			0	0	0	0	0	3
Music Online Reference ^{A1}	1	1	1	1	1	1	1	0	0	0	0	0	3
Smithsonian Global Sound ^{A3}	3	1	8	3	1	1		0	0	0	0	0	13
ARTstor ^{SW}							1	0	0	0	0	0	
CQ Press	18	1,615	260	1,138	1,138	2,117	2,947	0	0	0	0	0	10,000
CQ Researcher Plus Archive ^{SW}	18	1,615	260	1,138	1,138	2,117	2,947	0	0	0	0	0	20,000
Columbia University Press													0
Granger's World Poetry ^{SW}													



automating retrieval and reporting of database usage statistics for a consortium

- problem
 - what we were doing and why it doesn't work
- solution
 - the pieces, the parts and how they fit together
- future
 - what we've learned and our expectations

the solution

- automating
 - maintenance utilities
 - handling retrieved data
 - reporting in multiple formats
 - retrieval of vendor data

intranet web interface

Automated Vendor Statistics

[Queries](#)[Queue](#)[Utilities](#)[Users](#)[Reports](#)[Status](#)

Vendor Database Reports

The **College vs Period** page will display all colleges on the left side and periods across the top for

Calendar Year of or Fiscal Year of

for the vendor

the datatype of

the datasource of

The **Datasource vs Period** page will display all datasources on the left side and periods across the top for

Calendar Year of or Fiscal Year of

for the vendor

the datatype of

192.168.3.156x

Jim McGill

cclafla\jmcgill

Reference Code: /VendorDBMenu.asp

Automated Vendor Statistics

[Queries](#)[Queue](#)[Utilities](#)[Users](#)[Reports](#)[Status](#)

Vendor Database Utilities

Run The **Manual Entry** page is used to setup and enter data for

Run The **Add a new Vendor Name** for . This option will add a new Vendor record so that you may add a manual entry.

Run The **Download Documentation** page is used to change login documentation for . This information will be used to help update usernames, passwords, keys, etc. This option will also let you stop the automatic nightly download and to flag this vendor as Counter DB1 compliant.

Run The **Download Data** page is used to change login data for . This information will be used to help update usernames, passwords, keys, etc.

Run The **Vendor Links** page will allow you to request reports from vendors.

Run The **Vendor Display Names** for . This is used to assign standard college names to the Vendor college names.

Run The **Vendor Display Names** for all un-assigned vendors. This is used to assign standard college names to the Vendor college names that are not yet assigned.

Run Does **Data Exist** for the Calendar Year of or Fiscal Year of for ?

Run The **Update Datasources** Procedure will build a new DataSources Table. Run this if new datasources have been discovered.

Run The **Filter Type** page allows you to set up data for the filter type of .

Run The **Filter** page allows you to set up data for the filter type of and for the vendor .

Run The **College Data** page allows you to set up college data (size, region).

Run **Remove a DataSource** utility is used to delete a datasource from the statistics table. This is necessary when a Vendor changes the name of a database and the name change passes onto the reports. The system can not automatically know when this happens and will not overwrite the old records that use the original name. **DO NOT RUN THIS UNLESS YOU FULLY UNDERSTAND THE IMPLICATIONS.**

Delete Records for

Run The **Add a NEW Vendor** page is used to help add a new vendor. NOTE: You must be allowed to run this in menu/user security area.

192.168.3.156x

Jim McGill
ccfla\jmcgill

Reference Code: /VDBUtilities.asp

Automated Vendor Statistics

Queries

Queue

Utilities

Users

Reports

Status

VDB Last Run Status

	TOP=Hours since Last Run							MIDDLE=Run Duration in Seconds							BOTTOM=Latest Non-Zero Period in Database						
Brevard Community College	n/a	0,9 2	0,9 3	0,9 6	5,6 2	5,5 6	0,1 33 200801	3,8 5	3,8 13	3,7 6	3,7 12	3,6 14	3,5 18	3,4 18	n/a						
Broward Community College	n/a	n/a	0,9 2	0,9 4	5,6 2	5,5 5	0,1 29 200801	3,8 3	3,8 15	3,7 5	3,7 10	3,6 49 200801	3,5 10	3,3 6	n/a						
Central Florida Community College	n/a	n/a	0,9 2	5,6 8	5,5 2	5,5 6	0,0 31 200801	3,8 4	3,8 13	3,7 5	3,7 9	3,6 7	3,5 10	3,3 6	3,3 9						
Chipola College	n/a	n/a	0,9 2	5,6 3	5,5 1	5,5 4	0,0 17 200801	n/a	3,8 10	3,7 12	3,6 4	3,5 11	3,3 5	n/a							
Daytona Beach Community College	0,9 7 200712	n/a	0,9 2	5,6 4	5,5 2	5,5 5	0,0 22 200801	n/a	3,8 13	3,7 6	3,6 9	3,6 14	3,5 11	3,3 5	n/a						
Edison College	0,9 7 200712	n/a	0,9 3	5,6 3	5,5 2	5,5 5	0,0 25 200801	3,8 2	3,8 10	3,7 4	3,6 9	3,5 10	3,4 14	3,3 6	n/a						
Florida Community College at Jacksonville	n/a	n/a	0,9 2	5,6 4	5,5 2	5,5 7	0,0 27 200801	3,8 4	3,8 11	3,7 6	3,6 10	3,5 12	3,4 11	3,3 5	n/a						
Florida Keys Community College	n/a	n/a	0,9 2	5,6 3	5,5 1	5,5 5	0,0 -18157 200801	3,8 2	3,8 10	3,7 5	3,6 9	3,5 3	3,4 10	3,3 5	n/a						
Gulf Coast Community College	n/a	n/a	0,9 2	5,6 4	5,5 2	5,5 5	5,1 185 200801	3,8 2	3,7 10	3,7 4	3,6 10	3,5 7	3,4 10	3,3 5	n/a						

Vendor not responding

DataCheck for 2007

	Jan 2007	Feb 2007	Mar 2007	Apr 2007	May 2007	Jun 2007	Jul 2007	Aug 2007	Sep 2007	Oct 2007	Nov 2007	Dec 2007
Brevard Community College												
Broward Community College												
Central Florida Community College												
Chipola College												
College Center for Library Automation												
Daytona Beach Community College												
Edison College												
Florida Community College at Jacksonville												
Florida Keys Community College												
Gulf Coast Community College												
Hillsborough Community College												
Indian River Community College												
Lake City Community College												
Lake-Sumter Community College												
Manatee Community College												
Miami Dade College												
North Florida Community College												

Palm Beach Community College	▼	Palm Beach Community College
Palm Beach Community College	▼	pbcc.cc.fl.us
Pasco-Hernando Community College	▼	Pasco Hernando Community College
Pasco-Hernando Community College	▼	phcc.edu
Pensacola Junior College	▼	Pensacola Junior College
Pensacola Junior College	▼	pjc.edu
Polk Community College	▼	Polk Community College
Polk Community College	▼	polk.cc.fl.us
Santa Fe Community College	▼	Santa Fe Community College, Florida
Santa Fe Community College	▼	sfcc.edu
Seminole Community College	▼	scc-fl.edu
Seminole Community College	▼	Seminole Community College
South Florida Community College	▼	sfcc.cc.fl.us
South Florida Community College	▼	South Florida Community College
St. Johns River Community College	▼	sjrcc.edu
St. Johns River Community College	▼	St. Johns River Community College
St. Petersburg College	▼	spjc.edu
St. Petersburg College	▼	St. Petersburg Junior College
Tallahassee Community College	▼	Tallahassee Community College
Tallahassee Community College	▼	tcc.cc.fl.us
Valencia Community College	▼	Valencia Community College
Valencia Community College	▼	valenciacc.edu

Download Data for

Returns Calendar Year to Date

College Name is not used.

We use the "Ship To" account statistics.



Brevard Community College

College Name

BREVARD COMM COLL

UserName

XXXX

Password

XXXXXXXXXX

[Delete](#)

Broward Community College

College Name

BROWARD COMM COLL

UserName

XXXX

Password

XXXXXXXXXX

[Delete](#)

Central Florida Community College

College Name

CENTRAL FLORIDA COMM COLL

UserName

XXXX

Password

[Delete](#)

Automated Vendor Statistics

[Queries](#)
[Queue](#)
[Utilities](#)
[Users](#)
[Reports](#)
[Status](#)

Filter Rules for - Statewide Database

Data
Source

Start
Period

Stop
Period

Filter
Value

				Add Rule
Pick a DataSource	Earliest Period	Latest Period		+
Complete	200801	Latest Period		Delete
Premier	Earliest Period	Latest Period		Delete
views	Earliest Period	200712		Delete
Complete	200801	Latest Period		Delete
Premier	Earliest Period	Latest Period		Delete
	Earliest Period	Latest Period		Delete
tions	Earliest Period	Latest Period		Delete
rences	Earliest Period	Latest Period		Delete
Full Text	Earliest Period	Latest Period		Delete

College	Size Group	Region
Brevard Community College	Medium	Region 3
Broward Community College	Large	Region 5
Central Florida Community College	Small	Region 3
Chipola College	Small	Region 1
College Center for Library Automation	Other	Other
Daytona Beach Community College	Large	Region 1
Edison College	Medium	Region 4
Florida Community College at Jacksonville	Large	Region 2
Florida Keys Community College	Small	Region 5
Gulf Coast Community College	Medium	Region 1
Hillsborough Community College	Large	Region 4
Indian River Community College	Medium	Region 5
Lake City Community College	Small	Region 2
Lake-Sumter Community College	Small	Region 3
Manatee Community College	Medium	Region 4
Miami Dade College	Large	Region 5
North Florida Community College	Small	Region 2

Manual Entry

Change Scope

Searches Run

2007

Save

Delete

Exit

Vendor:

Data Source Display Name:

College	Jan 2007	Feb 2007	Mar 2007	Apr 2007	May 2007	Jun 2007	Jul 2007	Aug 2007	Sep 2007	Oct 2007	Nov 2007	Dec 2007	Ttl 2007
Brevard Community College	0	0	0	0	0	0	134	32	12	281	182	17	658
Broward Community College	0	0	0	0	0	0	429	2	56	267	477	120	1351
Central Florida Community College	0	0	0	0	0	0	55	10	119	211	74	3	472
Chipola College	0	0	0	0	0	0	0	0	28	14	18	0	60
College Center for Library Automation	0	0	0	0	0	0	0	0	0	0	0	0	0
Daytona Beach Community College	0	0	0	0	0	0	136	40	100	121	97	55	549
Edison College	0	0	0	0	0	0	12	53	8	11	239	11	334
Florida Community College at Jacksonville	0	0	0	0	0	0	34	11	42	48	300	64	499
Florida Keys Community College	0	0	0	0	0	0	8	11	0	0	1	0	20
Gulf Coast Community College	0	0	0	0	0	0	138	50	23	35	0	0	246

Automated Vendor Statistics

[Queries](#)[Queue](#)[Utilities](#)[Users](#)[Reports](#)[Status](#)

Maintain User Security

Run Edit **Existing User** Record for

Run Edit **Menu Security** Record for (# menu is not in list, enable security from the "edit" menu option below.)

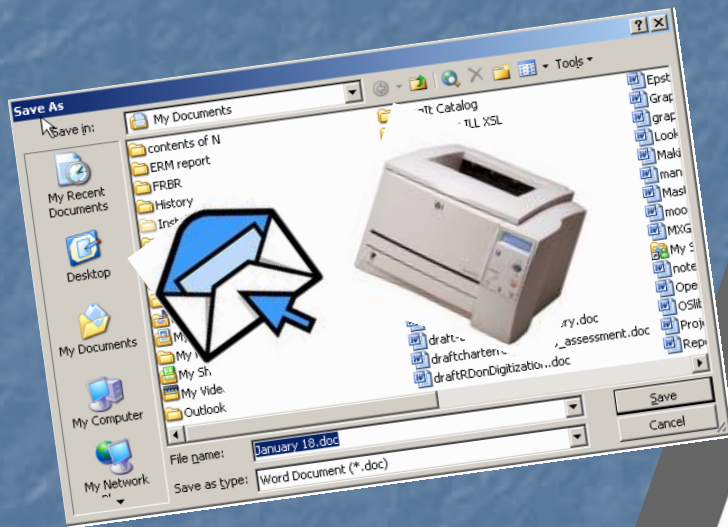
Run Add a **New User** Record for login name:

The **User List** for through . This report provides a summary of users accessing this system. This report will list users login name, first name, last name, time last logged into the system, number of times in the system and the last workstation address used. The names are linked so that records can be edited. Report Order:

- Run
- ☒ by most recent logged in
 - ☐ by login name alphabetical order
 - ☐ by last name alphabetical order
 - ☐ by most visits

Run Edit **Menu** record for

reporting





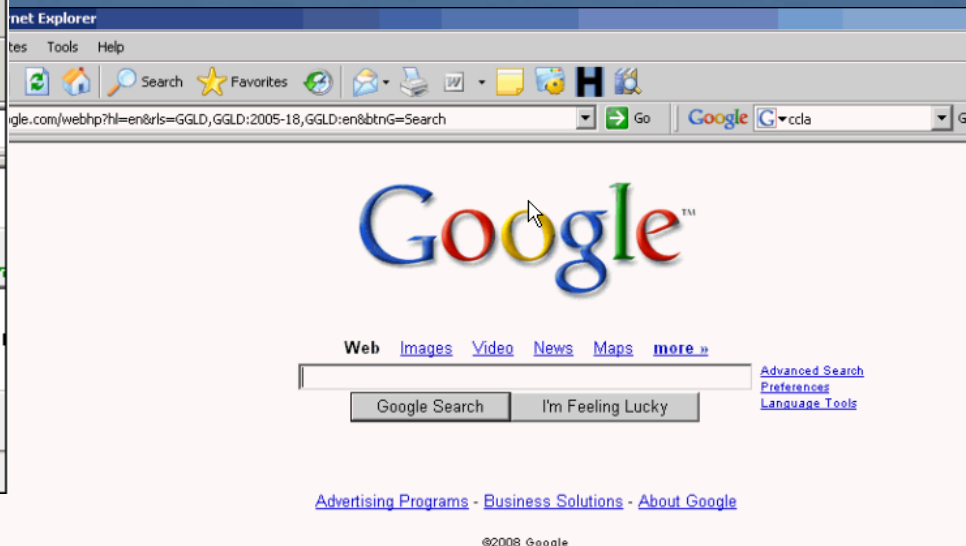
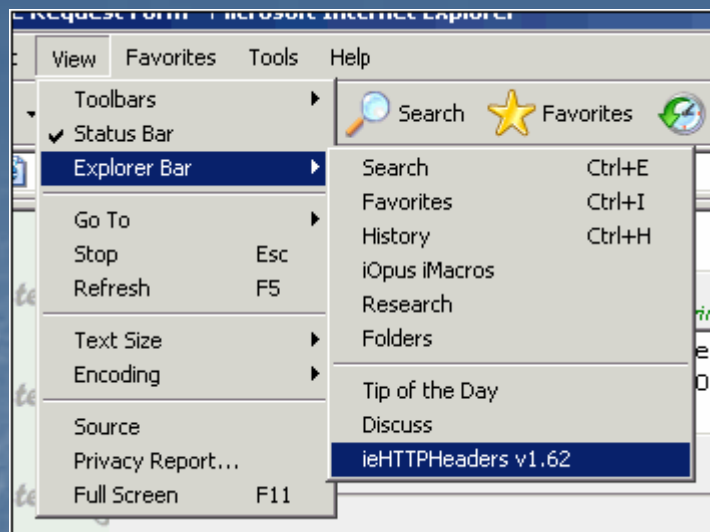
March 20, 2008

Electronic Resources and Libraries



creating retrieval scripts “nuts and bolts”





```
#####
# WebAutomaton.pl
# This script dynamically c
# Solution designed by: Dav
#
# Run this dynamically gene
# NOTE: This script only se
# response variables that y
# save it to a file or data
#
#####
```

```
# Use HTTP Web Client Libra
use LWP::UserAgent;
use HTTP::Request;
```

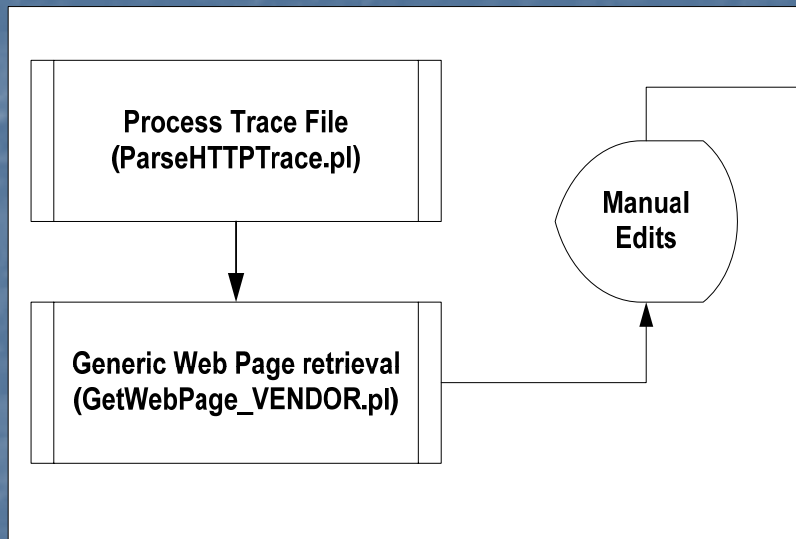
```
# create instance of user a
my $browser = LWP::UserAgen
# need to create temporary
# cookie stores the session ID used by websites
$browser->cookie_jar( {} );
```

```
#Make GET request to some URL with optional GET query parameters
my $resp0 = $browser->get("http://google.com");
# Response can be read like this:
# print $resp0->content;
```

```
Accept: image/gif, image/x-bitmap, image/jpeg, image/pjpeg, application/x-shockwave-flash, application/vnd.ms-excel, application/vnd.ms-powerpoint, application/msword, */*
Referer: http://www.google.com/search?hl=en&rls=GGLD,GGLD:2005-18,GGLD:en&btnG=Search
Accept-Language: en-us
Accept-Encoding: gzip, deflate
User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; .NET CLR 1.1.4322; .NET CLR 2.0.50727)
Host: www.google.com
Connection: Keep-Alive
Cookie: PREF=ID=de6f0cb6bc8334bb.TB-2.TM=1202911501.S=GKLNh_kF0gVS:s8je:NID=7=PDD0ys8SIzh4MaeIDrAgol_R9dUUW/m_hoCINa6V4AcD9_P3pt6dn/9x43Pm39kdTxdq2PbAAAn53dKlg/NmYgJwa%a6qKhcNwH3L_2h0YDvFp/SE4W6gV/C7Lq16tI

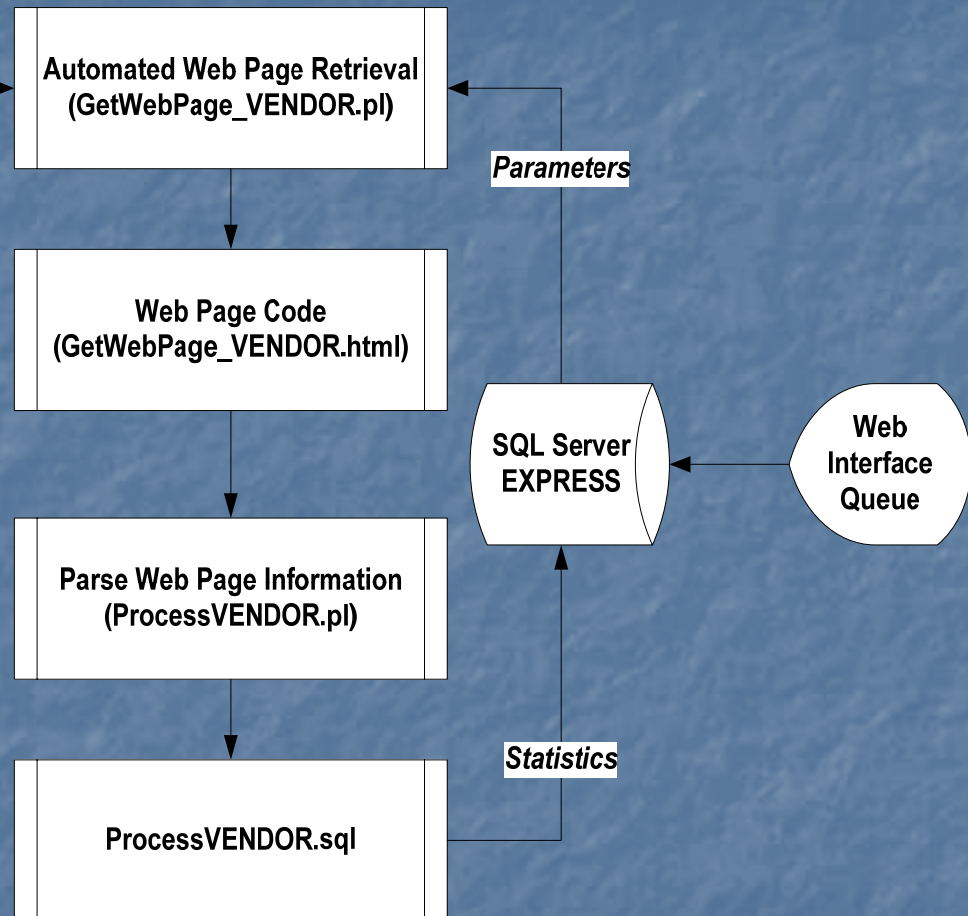
HTTP/1.1 200 OK
Cache-Control: private
Content-Type: text/html; charset=UTF-8
Content-Encoding: gzip
Server: gws
Content-Length: 2029
Date: Mon, 25 Feb 2008 13:25:25 GMT
```

One Time, 4 Step Process

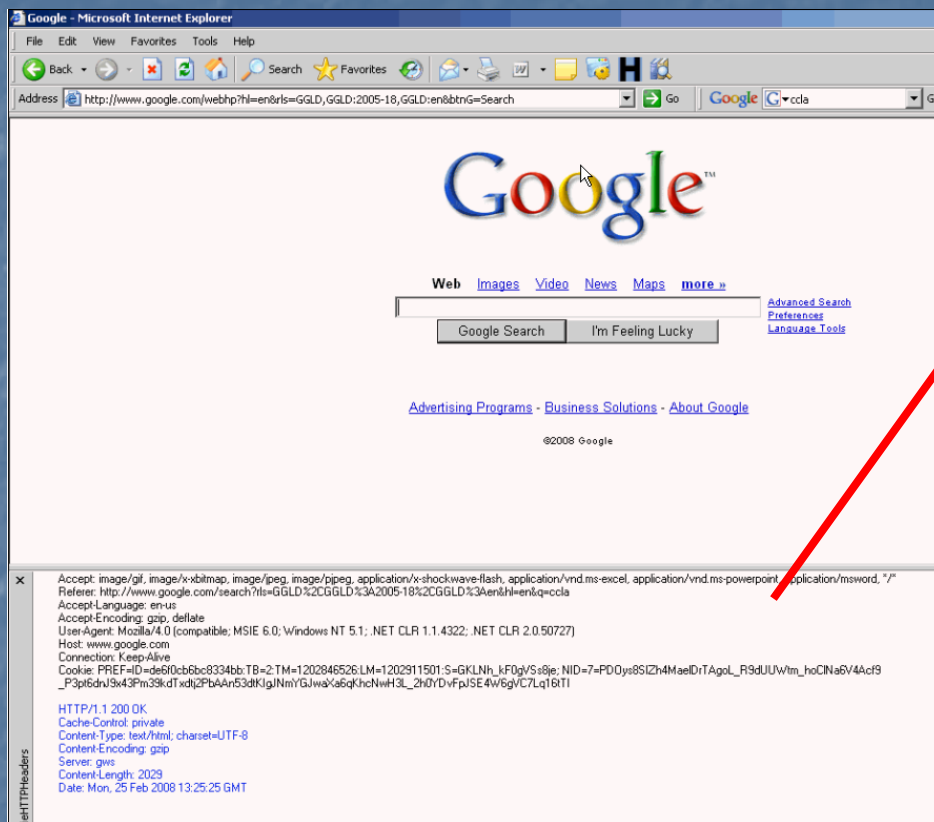


(Manual edits for testing & first cleanup – remove everything that isn't in table. This is iterative and run from the command prompt until satisfactory file is returned.)

Automated Process



step 1. capture HTTP headers







**Process Trace File
(ParseHTTPTrace.pl)**

**Generic Web Page retrieval
(GetWebPage_VENDOR.pl)**

This is a manual process to create the Perl script that will accept variables and create GetWebPage_VENDOR.pl

step 2. modify Perl script to accept command line variables

<code>\$Period=\$ARGV[0];</code>		YYYYMM - our DB format
<code>\$ScopeCustID=\$ARGV[1];</code>		vendor specific scope customer ID
<code>\$UserName=\$ARGV[2];</code>		
<code>\$Password=\$ARGV[3];</code>		
 <code>#\$ScopeCustID="bcc";</code>		remarks - unremarked for testing
<code>#\$Period="200701";</code>		
 <code>\$yr=substr(\$Period,0,4);</code>		to reformat standard YYYYMM format to two separate variables: MM and YYYY for URL
<code>\$mon=substr(\$Period,4,2);</code>		
<code>if (\$mon < 10) {\$mon=~s/0//gi};</code>		

**Automated Web Page Retrieval
(GetWebPage_VENDOR.pl)**

Step 3. modify script with command line variables and parse runtime variables

... **iodFromMonth=' . \$mon . '&timePeriodFromYear=' . \$yr . '&timeP ...**

```
$content0=$resp5->content;
```

```
$pos=index($content0,"VIEWSTATE")+13;
```

```
$pos2=substr($content0,$pos,5000);
```

```
$pos3=index($pos2,"value")+7;
```

```
$pos4=index($pos2,"\>");
```

```
$VIEWSTATE=substr($pos2,$pos3,$pos4-$pos3-2);
```

```
$VIEWSTATE=~s/\//\%2F/gi;
```

```
$VIEWSTATE=~s/\+/\%2B/gi;
```

```
$VIEWSTATE=~s/\=/\%3D/gi;
```

```
$pos=index($content0,"EVENTVALIDATION")+13;
```

```
$pos2=substr($content0,$pos,2000);
```

```
$pos3=index($pos2,"value")+7;
```

```
$pos4=index($pos2,"\>");
```

```
$EVENTVALIDATION=substr($pos2,$pos3,$pos4-$pos3-2);
```

```
$EVENTVALIDATION=~s/\//\%2F/gi;
```

```
$EVENTVALIDATION=~s/\+/\%2B/gi;
```

```
$EVENTVALIDATION=~s/\=/\%3D/gi;
```

SECURITY CODES

some codes are
session based &
must be parsed
out to pass to
subsequent pages

**Automated Web Page Retrieval
(GetWebPage_VENDOR.pl)**

step 4. create page parser (part 1)

creating ProcessVendor.pl script

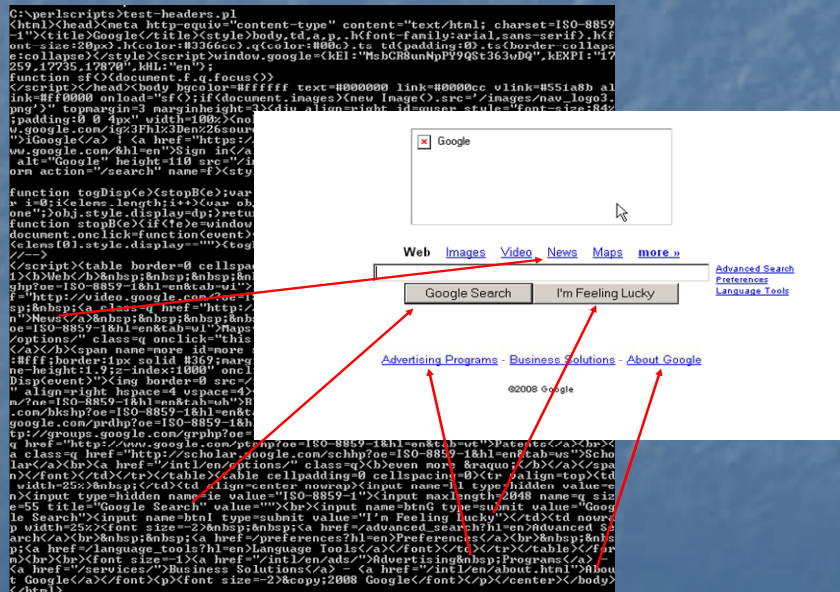
```
$col=$ARGV[0];  
$vendor="vendorname";  
$VDBSuffix="VENDOR";  
$jumpin="<b>Site:";  
$jumpout="Grand Total";  
require ("../VDBProcs.pl");
```

college name – when needed

anonymized (*for this presentation*) vendor name

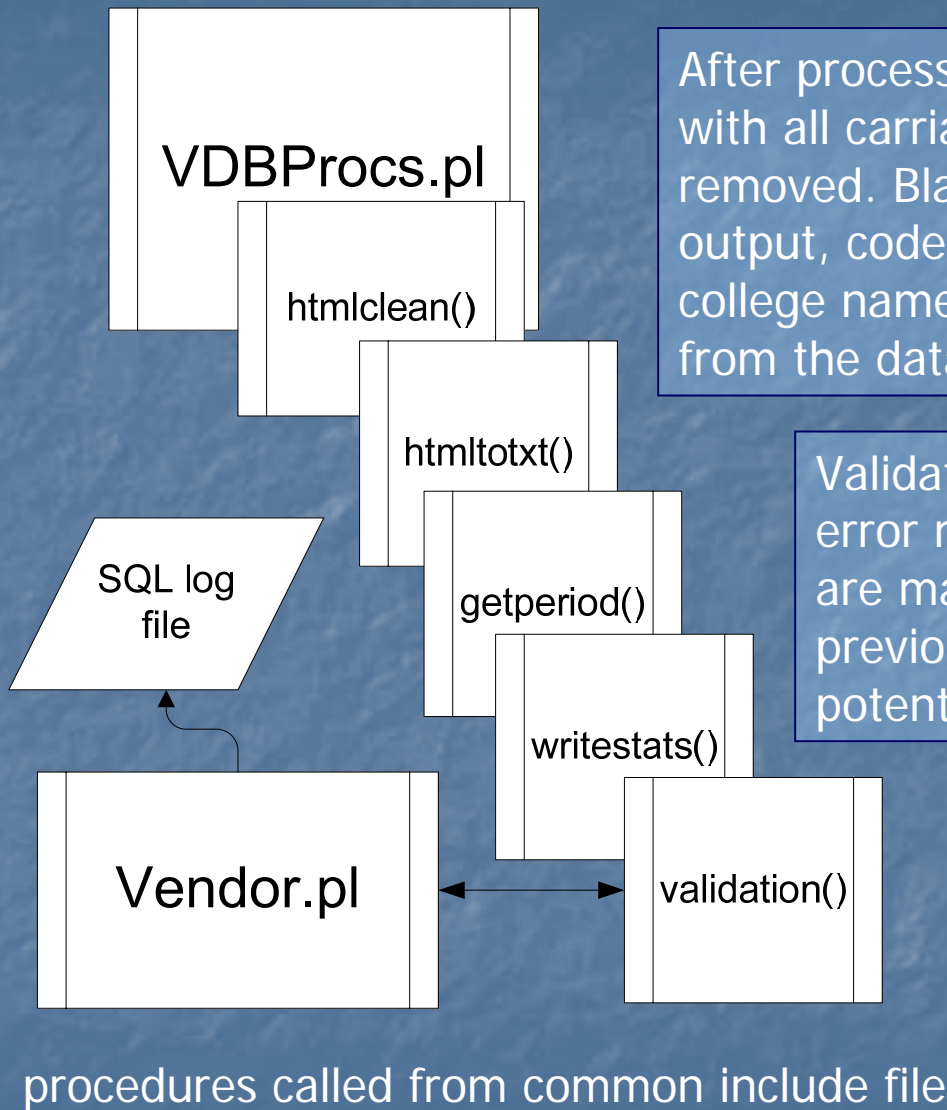
points to begin and stop processing file

include file with needed subroutines



**Parse Web Page Information
(ProcessVENDOR.pl)**

Step 4. create page parser (part 2)

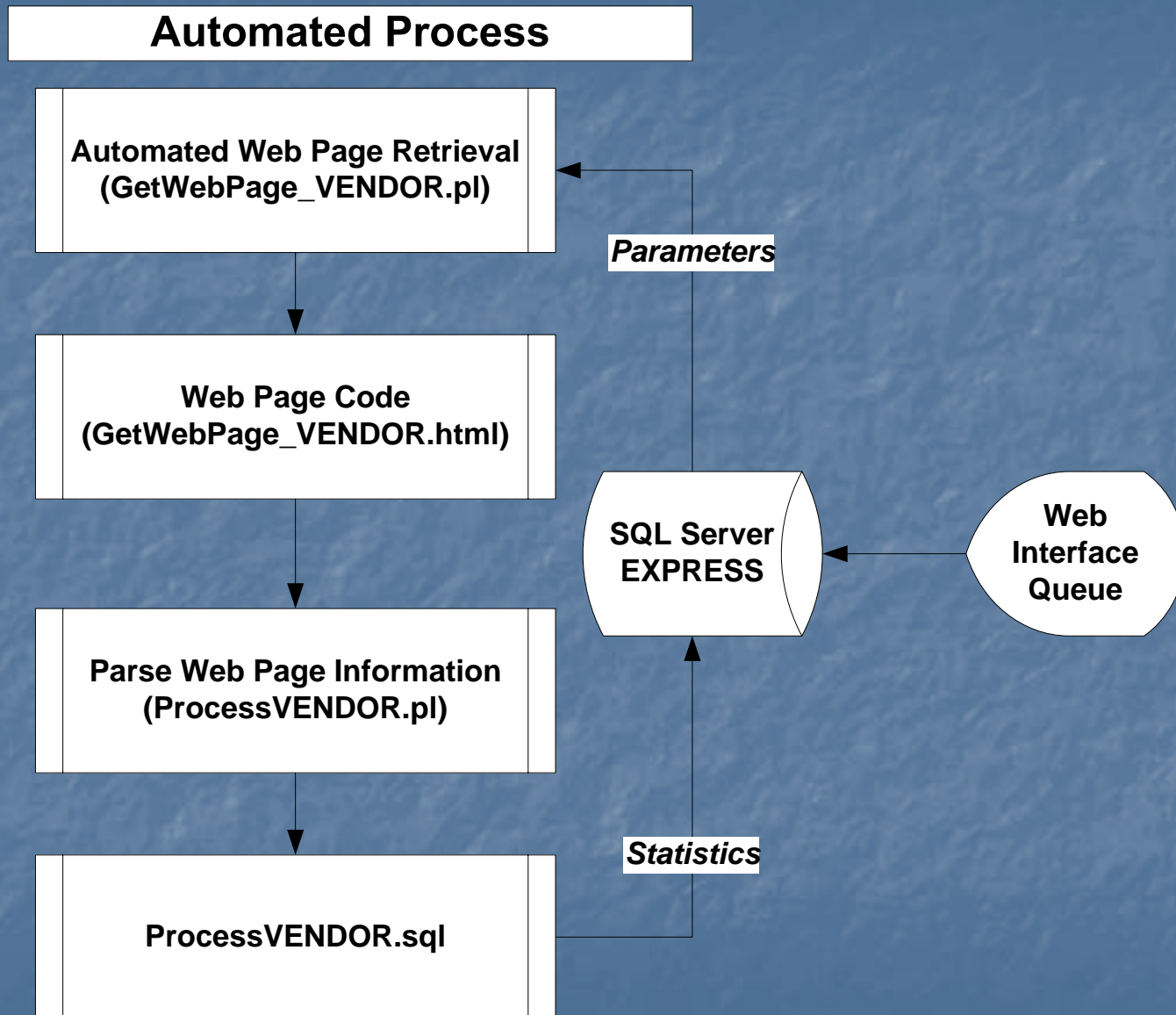


After processing, each table row is on one line with all carriage returns, linefeeds, and tabs removed. Blank lines and page feeds are not output, code outside jump* is ignored. Period, college name and other variables are passed from the database by the VDBProc.pl file.

Validation is run on SQL log file to look for error messages and write to log. Entries are made for no data, change from previously retrieved period value or other potential problems.

**Parse Web Page Information
(ProcessVENDOR.pl)**

automated process



handling retrieved data

```
delete from VDBStatistics where vendor='VENDOR'
and college='VALENCIA COMM COLLEGE'
and datasource='SOME VENDOR DATABASE'
and datatype='Sessions'
and subdatatype='0'
and period='200802'
insert into VDBStatistics
( sourcefile, vendor, college, period, datatype
, subdatatype, datasource, quantity )
values
('ProcessVENDOR.sql','VENDOR','VALENCIA COMM
COLLEGE'
,'200802','Sessions','0',SOME VENDOR
DATABASE','4348')
```

Table - dbo.VDBStatistics		Table - dbo.VDBStatistics*	Sum
	Column Name	Data Type	Allow Nulls
	Vendor	varchar(50)	<input type="checkbox"/>
	College	varchar(100)	<input type="checkbox"/>
	DataSource	varchar(200)	<input type="checkbox"/>
	DataType	varchar(100)	<input type="checkbox"/>
	SubDataType	varchar(100)	<input type="checkbox"/>
	Period	varchar(50)	<input type="checkbox"/>
	SourceFile	varchar(200)	<input checked="" type="checkbox"/>
	Quantity	int	<input checked="" type="checkbox"/>
	Cyear		<input checked="" type="checkbox"/>
	F1	varchar(3)	<input checked="" type="checkbox"/>
	F2	varchar(3)	<input checked="" type="checkbox"/>
	F3	varchar(3)	<input checked="" type="checkbox"/>
	F4	varchar(3)	<input checked="" type="checkbox"/>
	F5	varchar(3)	<input checked="" type="checkbox"/>

ProcessVENDOR.sql

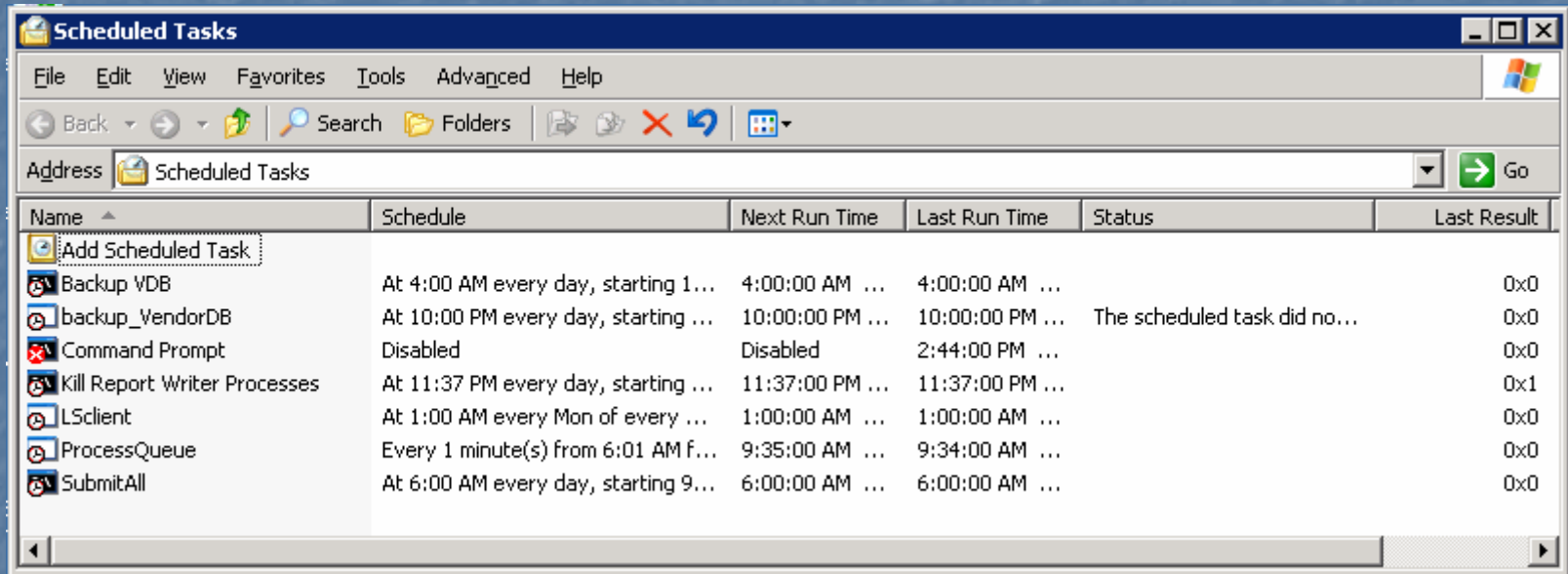
handling retrieved data

- where/how we store what we retrieve

Vendor	College	DataSource	DataType	Period	Qua...	Cyear
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200702	402	2006
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200702	705	2006
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200706	578	2006
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200707	452	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200707	375	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200708	90	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200708	133	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200709	287	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200709	765	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200710	745	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200710	981	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200711	1194	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200711	608	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200712	124	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200712	668	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200712	124	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200712	0	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200712	668	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200801	7	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200801	78	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200801	0	2007
VENDOR	Edison College	SOME VENDOR DATABASE	Searches Run	200801	133	2007

handling retrieved data

daily backup of database via windows scheduler



* SQL Server Express does not support SQL Agent

tools



software used

- retrieval of data – free
 - Internet Explorer
 - Perl
 - LWP library (Library for the WWW for Perl)
 - ieHTTP Headers
 - ParseHTTPTrace.pl
 - SQLExpress and manager
 - Intranet Site (IIS, .asp, vbscript, java)
- reporting – some cost
 - EZView (low cost)
 - Crystal Reports (had it)

structure

- environment
 - each vendor has its own working directory
 - each vendor has several files in this directory
 - batch file (called from SQL Server)
 - Perl script (gets web page)
 - Perl script (makes sql to load data)
 - log files (troubleshoot)

retrieval of vendor data

- activePerl 5.8.6 build 811 to download webpages
 - run from command prompt in development and testing
- ieHTTPHeaders - an add-on for IE that displays HTTP Headers
 - <http://www.blunck.se/iehttpheaders/iehttpheaders.html>
- once trace file is captured with ieHTTPHeaders add-on, use ParseHTTPTrace.pl to create GetWebPage_VENDOR.pl file.
 - <http://www.codeproject.com/KB/perl/webautomaton.aspx>

automating retrieval and reporting of database usage statistics for a consortium

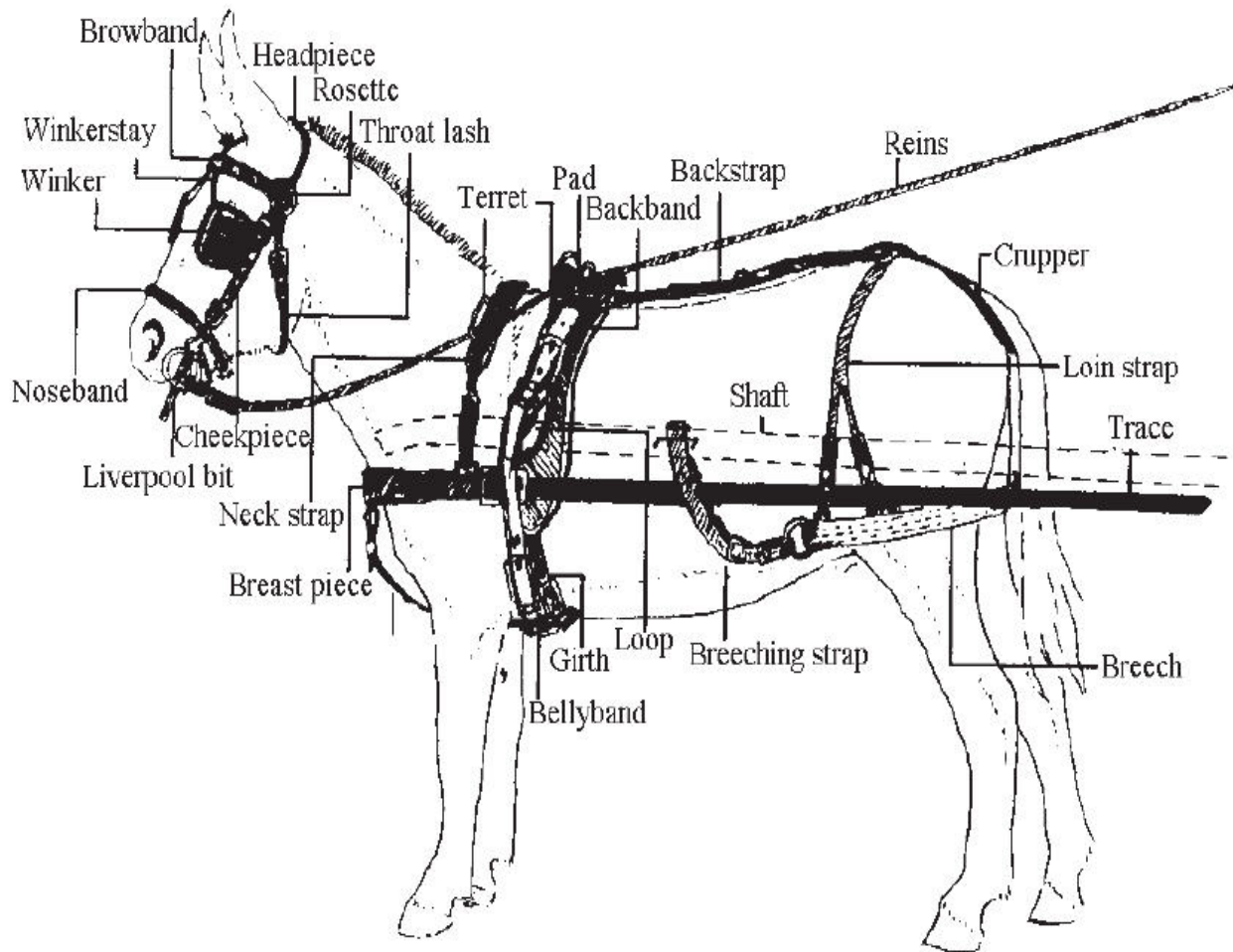
- problem
 - what we were doing and why it doesn't work
- solution
 - the pieces, the parts and how they fit together
- future
 - what we've learned and our expectations

what have we learned?

- large change in service requires staffing and support
- project name should be closely related to the service
- administration understanding of needs
 - assignment of priorities
 - proof-of-concept
 - need for ongoing support –vendor changes, local needs
- moving from proof-of-concept is NOT trivial
 - data checking/revisions/data checking/revisions
 - handoff from development to maintenance

expectations

- future use
 - until SUSHI is widespread OR
 - until data collection and reporting in ERM products is mature OR
 - until existing automated systems have reasonable consortial pricing
- future plans
 - customer/college interface
 - hope...



Thank you

College Center for Library Automation

1753 W. Paul Dirac Drive

Tallahassee, Florida 32310

Susan Campbell scampbell@cclaflorida.org

Jim McGill jmcgill@cclaflorida.org